36-303: Sampling, Surveys & Society
HW05: Hand in on Blackboard Thu Apr 06, 2012

Reminders:

- Please see Lecture 18, Week 09, for an outline of events for the rest of the semester.
- Reading:
  - Puetchel Handout on Oral Presenations (and another one on making attractive numerical tables!)
  - Bem Handout on Writing Papers
  - Examples of past 303 papers

  These are all under the link "Guides" near the "week09" link on http://www.stat.cmu.edu/~brian.

Exercises to Turn In (there are 4 exercises):

1. True/False:
   (a) To reduce the margin of error in a stratified sample, we should design strata so that the units within each stratum are as similar as possible.
   (b) To reduce the margin of error in a cluster sample, we should design clusters so that the units within each cluster are as similar as posslbe.
2. How would you stratify the population to produce smaller margins of error, for each of the following surveys? Name the stratifying variable(s) you would use for each survey.
   (a) A political poll to estimate the percentage of registered voters in Arizona who approve of the job the governor is doing.
   (b) A telephone survey of undergraduates at Carnegie Mellon, to estimate the total amount of money students spend on textbooks.
   (c) A sample of prime-time (8pm–11pm Monday through Saturday, and 7pm–11pm Sunday, in the US Eastern time zone) TV programs on CBS, to estimate the average number of promotional announcements (ads for other programs on the same network) per hour of broadcast.
3. Hard-shell clams can be sampled using a dredge (a large machie that scrapes material off the seabed in shallow waters). Clams do not tend to be uniformly distributed in a body of water however, because some areas provide better habitats than others, so taking an SRS is likely to result in a large estimated variance for the number of clams in the area. Russell (1972) used stratified random sampling to estimate the total number of bushels of hard-shell clams in Narragansett Bay, Rhode Island. The area of interest was divided into four strata based on preliminary surveys that identified areas in which clams were more, or less, abundant. Then, n_h dredge tows were made in each stratum h, for $h = 1, 2, 3, 4$. The acreage for each stratum was known, and Russell calculated that it would take 25.6 dredge tows to cover one whole acre of seabed.

[Continued on next page]

(a) Here are the results from the survey taken before the commercial fishing season. Estimate the total number of bushels of clams in the area, and give the standard error of your estimate. *Hint: First calculate N_h, the number of dredge tows needed to cover each stratum h, for $h = 1, 2, 3, 4$.*

Stratum	Area (Acres)	Number of Tows Made	Average Number of of Bushels per Tow	Sample Variance for Stratum
1	222.81	4	0.44	0.068
2	49.61	6	1.17	0.042
3	50.25	3	3.92	2.146
4	197.81	5	1.80	0.794

(b) The sample variance for an SRS of 18 dredge tows of the area represented by these four strata is estimated to be 1.084. What is the design effect in this case? Was it better to do a stratified sample than an SRS? By how much?

4. An inspector samples cans from a truckload of canned cream corn to estimate the average number of worm fragments per can. The truck contains 580 three-packs (three cans per package). The inspector samples 12 three-packs at random, opens all the cans, and finds the following number of worm fragments in each pack:

	Three-Pack											
Can	1	2	3	4	5	6	7	8	9	10	11	12
1	1	4	0	3	4	0	5	3	7	3	4	0
2	5	2	1	6	9	7	5	0	3	1	7	0
3	7	4	2	6	8	3	1	2	5	4	9	0

(a) Explain why this is a cluster sample and not a stratified sample.

(b) Estimate the mean number of worm fragments per can, along with the standard error of your estimate. *Hint: This follows very closely the example in the handout on clustered sampling.*

(c) Compute the design effect for this cluster sample.

(d) Estimate the *intraclass correlation*—that is, the correlation betwen worm fragment counts in cans from the same three-pack—using ideas from class.